



Computationally efficient and robust frequency-domain GSC

Ludovick Lepauloux, Pascal Scalart, Claude Marro

► To cite this version:

Ludovick Lepauloux, Pascal Scalart, Claude Marro. Computationally efficient and robust frequency-domain GSC. 12th IEEE International Workshop on Acoustic Echo and Noise Control, Aug 2010, Tel-Aviv, Israel. inria-00512654

HAL Id: inria-00512654

<https://inria.hal.science/inria-00512654>

Submitted on 31 Aug 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Computationally efficient and robust frequency-domain GSC

Ludovick Lepauloux Orange Labs - TECH/OPERA

22307 Lannion Cedex, FRANCE

Email: ludovick.lepauloux@me.com

Pascal Scalart ENSSAT - IRISA/CAIRN

22305 Lannion Cedex, FRANCE

Email: pascal.scalart@enssat.fr

Claude Marro Orange Labs - TECH/OPERA

22307 Lannion Cedex, FRANCE

Email: claudemarro@orange-ftgroup.com

Abstract—In this paper, we propose a novel adaptive beamforming algorithm with low signal distortion capability and high computational efficiency. The proposed algorithm incorporates the crosstalk-resistant adaptive noise canceller (CTRANC) in place of the usual adaptive interference canceller (AIC). Doing so, a better signal leakage estimate is provided at the adaptive interference canceller input but with lower computational cost. Consequently, the noise and interference adaptation process is improved which further minimizes the signal cancellation at the output. Based on the robust generalized sidelobe canceller (RGSC), the proposed algorithm moreover guarantees robustness to steering vector errors.

I. INTRODUCTION

A robust and computationally efficient adaptive beamforming algorithm is an essential signal processing tool for enhancing a speech signal received by a microphones array in a noisy and reverberant environment. Beamformers based on generalized sidelobe cancellers (GSC) [1] have been extensively studied in the literature. It consists of a presteering front end, a fixed beamformer, a blocking matrix and an adaptive canceller. The presteering front end is composed of variable time delays allowing the main lobe of the beamformer to be steered to the desired direction. The fixed beamformer (FBF) is used to enhance the target-signal from the look direction. The blocking matrix (BM), composed with adaptive blocking filters (ABF), rejects the target-signal, so that the blocking matrix output contains interference and noise. The adaptive canceller, composed with adaptive canceling filters (ACF), is able to adjust its weights so that the interferences and noise can be subtracted from the fixed beamformer output. However, the classical adaptive beamformer based on the GSC, like the simple Griffiths-Jim beamformer (GJBF) [1] suffers from target-signal cancellation due to steering-vector errors. These steering-vector errors are due to errors in microphone positions, microphone gains and real world recordings (reverberation, noise and moving target). Indeed, the beamformer is constrained to produce a dominant response toward the desired speech source location, while minimizing the response in all other directions. However, in reverberant environments a single direction of arrival cannot be determined since the desired signal and its reflexion impinge on the array from several directions. Thus, complete rejection of the target-signal is almost impossible in the BM and a considerable portion of the desired speech will leak to the interference canceller which results in target-signal cancellation.

To reduce target-signal cancellation, a variety of techniques have been proposed. Jan and Flanagan [2] suggested a matched filter beamforming (MFBF) instead of the conventional delay and sum beamformer (DSBF). This method has been further analyzed and improved by Gannot [3] and later by Reuven [4], [5]. Derivative constraints [6], [7], [8] and leakage constraints [9] to broaden the width of each adapted beam or target tracking [10] and calibration

[11] are other approaches for robust beamforming. Nevertheless, these latter require a large number of linear constraints or computations cost increase, which is in contradiction with our goal to define a robust computationally efficient frequency-domain GSC.

Recently, we have proposed an algorithmic and memory complexity reduction of the crosstalk-resistant adaptive noise canceller (CTRANC) [12]. In this paper, we propose to take benefit of the ability of this structure to deal with crosstalk problem which in our case is the same as the signal leakage problem in the GSC. The closest existing work in the literature motivated by the same observations is given by Choi et al. in [13] and described in Fig. 1. Here, we propose to use the CTRANC approach in place of the adaptive interference canceller (AIC) used in the RGSC defined by Hoshuyama et al. in [14] and shown in Fig. 2. This one uses an adaptive blocking matrix (ABM) with coefficients-constrained adaptive filters, which prevents the target-signal from leaking into the AIC. In addition to this, the AIC uses norm-constrained adaptive filters that can further improve the robustness against target-signal cancellation. Optimal implementation in the frequency-domain has been proposed by Herbordt and Kellerman in [15].

In Section II-C, we describe the derivation of the CTRANC algorithm in the frequency-domain and obtained thanks to our recent work [12]. The proposed robust GSC is also given. Section III describes the experimental results and we conclude and outline our future work in Section IV.

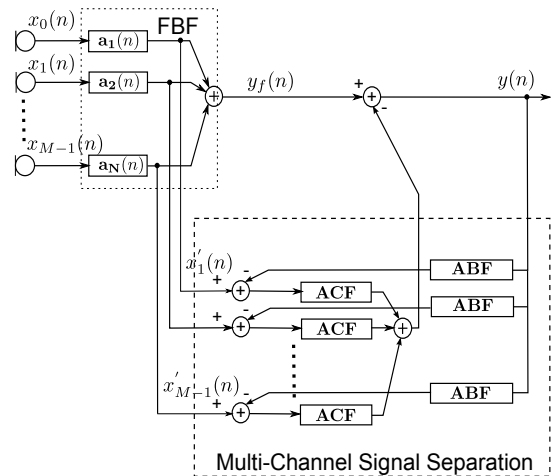


Fig. 1. Robust time-domain GSC having its ABF and ACF connected in feedback.

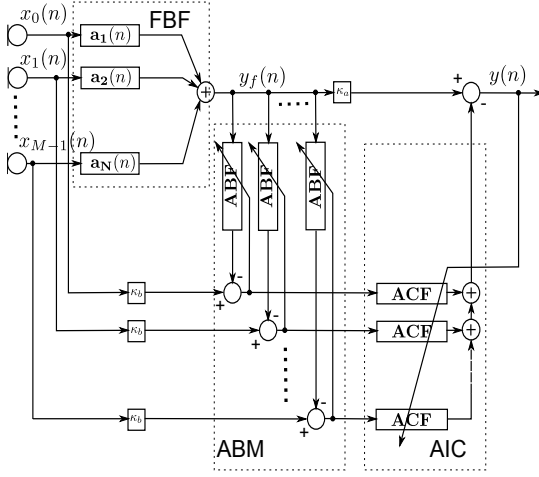


Fig. 2. Robust time-domain GSC.

II. FREQUENCY-DOMAIN CTRANC AND ROBUST GSC

In the following, uppercase symbols denote variables in the frequency domain, lowercase symbols stand for time-domain variables, and the boldface font indicates a vector or matrix quantity. Superscripts T and H represent transpose and complex conjugate transpose, respectively. The number of microphones is denoted by M , the discrete Fourier transform (DFT) length is $2N$. \mathbf{F} is the $2N \times 2N$ DFT matrix. The discrete time variable is denoted by n . We further use the time index k that reflects the discrete time in numbers of blocks.

A. Frequency-domain CTRANC (FCTRANC)

The CTRANC structure is given in Fig.4. In [12], we have proved that recursive gradient formula in CTRANC algorithm, described in [16] and [17], could be rewritten as convolution operation without a loss in performance. Thanks to this optimization, we have drastically reduced the memory cost and algorithmic complexity and made direct frequency-domain implementation possible. Algorithmic description is done as for the FLMS gradient block described in [18] and is given here after. However, the main difference holds in the block size. Indeed, our algorithm computes the convolution and the correlation of 3 vectors. Consequently, in order to avoid time aliasing, we must use a block size of $N = 2L$ samples rather than $N = L$ as done classically. Secondly, note that in the described algorithm the gradient quantities are limited to N correct samples, which corresponds to a truncated version of the theoretical ones. This is related to the observations made in [12] where we have shown the necessity to constraint the length of the recursion of the gradient formula with speech signals, due to potential instabilities. In addition, it is possible to prove analytically that the more significant values are contained in this first N samples.

B. Step-size control

In order to be compliant with the non-stationarities of the speech signal, we propose to replace the constant step-size 2μ in the described algorithm by two adaptive step-sizes. These are obtained thanks to the two-step noise reduction technique [19]. Indeed, we define the two following adaptive step-sizes:

$$\begin{aligned} \mu_{\mathbf{W}^1}(k) &= \frac{1}{1 + \text{SNR}_{prio}(k)} \\ \mu_{\mathbf{W}^2}(k) &= \frac{\text{SNR}_{prio}(k)}{1 + \text{SNR}_{prio}(k)} \end{aligned}$$

FCTRANC ALGORITHM BASED ON OVERLAP-SAVE SECTIONING

INITIALIZATION

$$\begin{aligned} \mathbf{W}^1(0) &= [0, \dots, 0]^T & \text{GradE1old}(0) &= [0, \dots, 0]^T & \mathbf{e}_1(0) &= [0_N]^T \\ \mathbf{W}^2(0) &= [0, \dots, 0]^T & \text{GradE2old}(0) &= [0, \dots, 0]^T & \\ \mathbf{G}(0) &= [0, \dots, 0]^T & Pr_m(0) &= Pf_m(0) = \delta_m & m &= 0, \dots, 2N-1 \end{aligned}$$

MATRIX DEFINITIONS:

$$\mathbf{g} = \begin{bmatrix} \mathbf{I}_N & \mathbf{0}_N \\ \mathbf{0}_N & \mathbf{0}_N \end{bmatrix} \quad \mathbf{k} = [0_N \mathbf{I}_N]; \text{ sectioning constraints}$$

$$\mathbf{F} = 2N \times 2N \text{ DFT matrix}$$

FOR EACH NEW BLOCK OF N INPUT SAMPLES:

Signal from the ABM output: Ob(k)

$$\mathbf{X}_r(k) = \text{diag}\{\mathbf{F}[e_1(kN-N), \dots, e_1(kN-1), e_1(kN), \dots, e_1(kN+N-1)]^T\}$$

Convolution

$$\mathbf{Y}_2(k) = \mathbf{X}_r(k) \mathbf{W}^2(k)$$

OLS

$$\mathbf{y}_2(k) = \mathbf{k} \mathbf{F}^{-1} \mathbf{Y}_2(k)$$

Error calculus

$$\mathbf{ob}(k) = \mathbf{k} \mathbf{F}^{-1} \mathbf{Ob}(k)$$

$$\mathbf{e}_2(k) = \mathbf{ob}(k) - \mathbf{y}_2(k)$$

$$\mathbf{E}_2(k) = \mathbf{F} \mathbf{k}^T \mathbf{e}_2(k)$$

Step size

$$Pr_m(k) = \lambda Pr_m(k-1) + (1-\lambda) |X_{r_m}(k)|^2 \quad m = 0, \dots, 2N-1$$

$$\mu_r(k) = \text{diag}\{Pr_0^{-1}(k), \dots, Pr_{2N-1}^{-1}(k)\}$$

Stochastic gradient

$$\mathbf{G}(k) = \mathbf{W}^1(k) \mathbf{W}^2(k)$$

$$\mathbf{W}^2(k+1) = \mathbf{W}^2(k) + 2\mu \mathbf{F} \mathbf{g} \mathbf{F}^{-1} \mu_r(k) \mathbf{E}_2(k) [\mathbf{X}_r^H(k) - (\mathbf{G}(k) \text{GradE2old}(k))^H]$$

$$\text{GradE2old}(k) = \mathbf{F} \mathbf{g} \mathbf{F}^{-1} \mu_r(k) \mathbf{E}_2(k) [\mathbf{X}_r^H(k) - (\mathbf{G}(k) \text{GradE2old}(k))^H]$$

Signal from the Fixed Beamformer output: y_f(n)

$$\mathbf{X}_f(k) = \text{diag}\{\mathbf{F}[e_2(kN-N), \dots, e_2(kN-1), e_2(kN), \dots, e_2(kN+N-1)]^T\}$$

Convolution

$$\mathbf{Y}_1(k) = \mathbf{X}_f(k) \mathbf{W}^1(k)$$

OLS

$$\mathbf{y}_1(k) = \mathbf{k} \mathbf{F}^{-1} \mathbf{Y}_1(k)$$

Error calculus

$$\mathbf{e}_1(k) = \mathbf{y}_f(k) - \mathbf{y}_1(k)$$

$$\mathbf{E}_1(k) = \mathbf{F} \mathbf{k}^T \mathbf{e}_1(k)$$

Step size

$$Pf_m(k) = \lambda Pf_m(k-1) + (1-\lambda) |X_{f_m}(k)|^2 \quad m = 0, \dots, 2N-1$$

$$\mu_f(k) = \text{diag}\{Pf_0^{-1}(k), \dots, Pf_{2N-1}^{-1}(k)\}$$

Stochastic gradient

$$\mathbf{W}^1(k+1) = \mathbf{W}^1(k) + 2\mu \mathbf{F} \mathbf{g} \mathbf{F}^{-1} \mu_f(k) \mathbf{E}_1(k) [\mathbf{X}_f^H(k) - (\mathbf{G}(k) \text{GradE1old}(k))^H]$$

$$\text{GradE1old}(k) = \mathbf{F} \mathbf{g} \mathbf{F}^{-1} \mu_f(k) \mathbf{E}_1(k) [\mathbf{X}_f^H(k) - (\mathbf{G}(k) \text{GradE1old}(k))^H]$$

where SNR_{prio} stands for the *a priori* signal-to-noise ratio (SNR) measured on $y_f(n)$. According to these expressions, each step-size evolves in the interval $[0,1]$ and are complementary. Consequently, these adaptive step-sizes ensure that each filter is adapted at the right instant. As an example, Fig.3 shows the behavior of the adaptive step-sizes for an input SNR of 15 dB. We can observe that these ones are effectively complementary in each frequency bin, which enables to drive accurately the adaptation. Note that each method is evaluated with these step-sizes.

C. Proposed robust GSC

The complete algorithm for the proposed robust GSC is obtained by connecting directly the CTRANC algorithm to the fixed beamformer and ABM outputs defined in [15] and shown in Fig. 2. These one are kept unchanged. The proposed robust GSC is shown in Fig. 4.

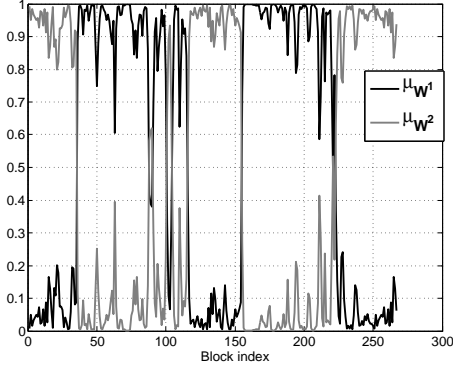


Fig. 3. Behavior of the proposed adaptive step-sizes for the frequency bin $k = 1 \text{ kHz}$.

As it can be observed, the recursive structure which stands for the AIC is only made with two adaptive filters. Simulation results will prove that in spite of this complexity reduction, the signal distortion is reduced in comparison with the one proposed by Hoshuyama et al. This can be explained by the filter \mathbf{W}^2 that further suppresses the target-signal leakage as it is done firstly by the ABM.

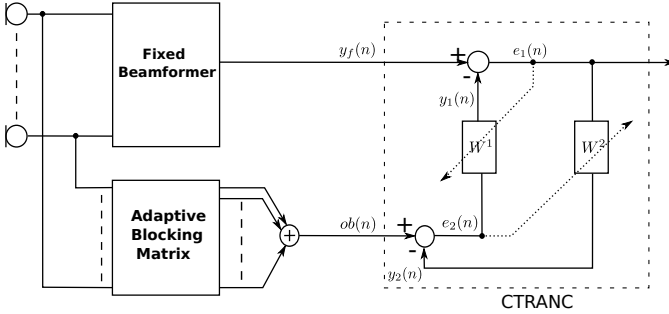


Fig. 4. Block diagram of the proposed robust GSC.

III. EXPERIMENTAL RESULTS

Four different implementations of the frequency-domain adaptive GSC are compared. The robust GSC given by Herbordt in [15] and based on the one proposed by Hoshuyama. The recursive implementation proposed by Choi et al. [13] and shown in Fig. 1. The one proposed and a modified version of the structure given by Choi where we use the proposed algorithm in place of the information maximization learning rules. They have been evaluated in a real videoconferencing context. In this environment, the perturbation generated by the acoustical environment is composed of the reverberation and a localized noise source. This section describes the experimental procedure and presents the results.

A. Experimentation procedure

1) *Array implementation:* In order to avoid spectral distortions of the output signal and to be less sensitive to steering errors, we use a directivity-controlled array. It is made of eleven cardioid microphones clustered in four subarrays. This enable us to ensure constant directivity over the frequency band [300 8000] Hz. The front end geometry and characteristics are fully defined in [20]. An exact localization of the desired source is assumed. The window length includes 512 samples (32 ms at a sampling rate of 16 kHz) with an overlap of 256 samples for the feedforward implementation. In case

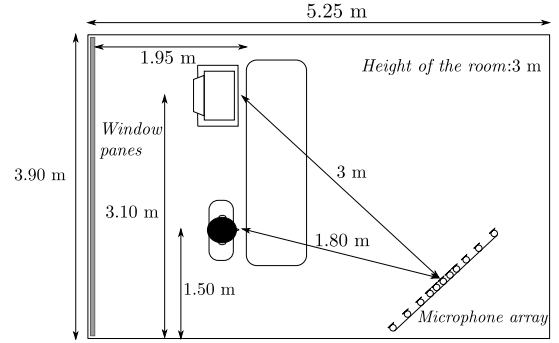


Fig. 5. Configuration of source/receiver location in the room.

of the feedback versions, we use respectively 256 and 128 samples. The analysis and synthesis windows have the perfect reconstruction property. Time aliasing is avoided by zero padding the block of samples before the fast Fourier transform as it is done in the frequency LMS algorithm (FLMS) with OLS technique [21]. The spectral densities (PSD) are estimated by means of an exponentially weighted averaging procedure. The time constant is equal to 64 ms which has been found to be a compromise between a low variance of the estimator and a fast updating. Consequently, the averaging factor in the FCTRANC is chosen such that $\lambda = 0.77$.

2) *Experimental setup:* The performance of the studied methods was measured in the real videoconferencing context described in [20]. For this purpose, impulse responses were measured in the room. The dimensions of this room are $5.25 \times 3.9 \times 3 \text{ m}$ and its reverberation time varies from 650 ms at 125 kHz to 250 ms at 4 kHz. The microphone signals are obtained by filtering the anechoic signals by the measured impulse responses between two loudspeakers and the eleven sensors of the array. The arrangement of the two sound sources with respect to the array are shown in Fig. 5. The input SNRs (SNR_{in}) are computed using the ITU-T recommendation P.56 speech voltmeter (SV56). Office noise was used as a noise source, and four english sentences (two male voices and two female voices) as source signals with a sample rate of 16 kHz.

3) *Objective measurements:* To evaluate the studied systems, we propose to measure the noise reduction factor, the SNR gain and the distortion of the desired signal. The noise reduction factor (**NR**) is measured during speech absence whereas the SNR gain (**G**) is measured during speech presence. For these two criteria a three-step procedure as described in [5] is used. First, each adaptive filter is computed from the microphone signals and stored in each frame. Then, they are applied to the noise alone and finally to the source signal alone. The stored adaptive filters are also applied to the signal obtained by filtering the anechoic signal by the direct path of the measured impulse responses. The distortion is given by the cepstral distance (**CD**) between the reference signal and the corresponding output signal, as done in [20]. The reference signal is defined as the filtering of the anechoic signal by the direct path associated to the central microphone. The cepstral distance is measured over consecutive time segments of 16 ms.

B. Results

For the results given hereafter, note that an ideal vocal activity detector is used for the estimation of the noise PSD. Table I summarizes the **NR**, **G** and **CD** values. As expected, our proposed robust GSC offers less noise reduction in comparison with the other methods. This can be explained by the fact that only one filter is used for producing a noise estimate, which is less effective especially in case of low SNRs. However, the cepstral distance achieved by the proposed robust GSC

	NR [dB]				G [dB]				CD			
Herbordt	16.2	13.9	11.9	10.8	11.8	10.8	9.5	8.6	0.54	0.53	0.5	0.47
Proposed	14.6	12.1	9.8	8.5	9.1	8	6.5	5.6	0.44	0.42	0.42	0.41
Choi	16.7	13.4	12.3	13.2	6.8	5.6	5.4	5.8	0.41	0.38	0.36	0.36
Choi (modified version)	14.7	15	14.5	13.5	8.7	9	8.7	8.1	0.38	0.38	0.36	0.39
SNR _{in}	0	5	10	15	0	5	10	15	0	5	10	15

TABLE I
NOISE REDUCTION (NR), SNR GAIN (G) AND CEPSTRAL DISTANCE (CD) MEASURED IN OFFICE NOISE.

is always better than the one achieved by the direct structure given by Herbordt in [15]. This observation shows that the second filter \mathbf{W}^2 connected recursively clearly improves the signal quality. This finding is further reinforced by noting that the best cepstral distance values are obtained with the recursive implementations of the GSC whatever the used algorithm. Nevertheless, best results are obtained with the modified version of the structure proposed by Choi et al., combining the adaptive step-sizes with the frequency-domain implementation of the optimized CTRANC algorithm [12] and described in subsection II-A. This can be firstly observed by comparing the SNR gain \mathbf{G} obtained with the information maximization learning rules [Choi] and the proposed method. We can see that the proposed structure is slightly better than this last one from 2 dB in case of input SNRs equal 0 and 5 dB and 1 dB for an input SNR of 10 dB, in spite of its low complexity regarding the number of filters used. Then, if we take into account the three criteria without considering the complexity, the more effective method corresponds to the recursive implementation of the GSC given by Choi but with our algorithm in place of the information maximization learning rules [Choi (modified version)]. This is due to the fact that our algorithm better takes into account the cross-coupling effect of the CTRANC approach, as it was already explained in [12]. These results are corroborated by informal listening tests.

IV. CONCLUSION

In this paper, we have proposed a novel recursive structure for the GSC based on the RGSC architecture, and with its appropriate adaptive algorithm. Doing so, we firstly ensure robustness against steering vector errors. Then, by taking benefit from the recursive structure (CTRANC), the proposed beamformer reduces the number of computations for the noise cancellation stage. And in spite of this optimization, we have shown that the signal distortion is reduced but at the expense of less noise reduction in comparison with the feedforward implementation of the GSC given by Hoshuyama et al. However, in order to emphasize the advantages of an AIC and ABF connected in feedback, a second recursive implementation proposed by Choi et al. but updated with our optimized algorithm in the frequency-domain is also given. Experimental results show that the proposed approach, based on a modified version of the structure given by Choi et al., is the most effective of all the considered methods.

REFERENCES

- [1] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," in *IEEE Transactions on Antennas and Propagation*, vol. 30, no. 1, January 1982, pp. 27–34.
- [2] E.-E. Jan and J. Flanagan, "Sound capture from spatial volumes: Matched-filter processing of microphone arrays having randomly-distributed sensors," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Atlanta, 1996, pp. 917–920.
- [3] S. Gannot, D. Burnstein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1614–1626, August 2001.
- [4] G. Reuven, S. Gannot, and I. Cohen, "Performance analysis of dual source transfer-function generalized sidelobe canceller," *Speech Communication*, vol. 49, no. 7–8, pp. 602–622, July–August 2007.
- [5] —, "Dual-source transfer-function generalized sidelobe canceller," *IEEE Transaction on Audio, Speech and Language Processing*, vol. 16, no. 4, pp. 711–727, May 2008.
- [6] M. H. Er and A. Cantoni, "Derivative constraints for broad-band element space antenna array processors," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 31, no. 6, pp. 1378–1393, December 1983.
- [7] K. M. Buckley and L. J. Griffiths, "An adaptive generalized sidelobe canceller with derivative constraints," *IEEE Transactions on antennas and propagation*, vol. 34, no. 3, pp. 311–319, March 1986.
- [8] G. L. Fudge and D. A. Linebarger, "Spatial blocking filter derivative constraints for the generalized sidelobe canceller and music," *IEEE Transactions on Signal Processing*, vol. 44, no. 1, pp. 51–61, January 1996.
- [9] Q. Z. Yongzhi Liu and Z. Lin, "Generalized sidelobe cancellers with leakage constraints," in *IEEE International Symposium on Circuits and System*, vol. 4, May 2005, pp. 3741–3744.
- [10] S. Affes and Y. Grenier, "A source subspace tracking array of microphones for double talk situations," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, May 1996, pp. 909–912.
- [11] G. L. Fudge and D. A. Linebarger, "A calibrated generalized sidelobe canceller for wideband beamforming," *IEEE Transactions on Signal Processing*, vol. 42, no. 10, pp. 2871–2875, October 1994.
- [12] L. Lepauloux, P. Scalart, and C. Marro, "An efficient low-complexity algorithm for crosstalk-resistant adaptive noise canceller," in *European Signal Processing Conference*, August 2009, pp. 204–208.
- [13] C. Choi, D. Kong, J. Kim, and S. Bang, "Speech enhancement and recognition using circular microphone array for service robots," in *International Conference on Intelligent Robots and Systems*, Las Vegas, October 2003, pp. 3516–3521.
- [14] O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," in *IEEE Transactions on Signal Processing*, vol. 47, no. 10, October 1999, pp. 2677–2684.
- [15] W. Herbordt and W. Kellermann, "Efficient frequency-domain realization of robust generalized sidelobe cancellers," in *IEEE fourth Workshop on Multimedia Signal Processing*, October 2001, pp. 377–382.
- [16] R. L. Zinser and J. B. Evans, "Some experimental and theoretical results using a new adaptive filter structure for noise cancellation in the presence of crosstalk," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 10, April 1985, pp. 1253–1256.
- [17] G. Mirchandani, R. L. Zinser, Jr., and J. B. Evans, "A new adaptive noise cancellation scheme in the presence of crosstalk," *IEEE Transactions on circuits and systems II: Analog and Digital Signal Processing*, vol. 39, no. 10, pp. 681–694, October 1992.
- [18] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Processing Magazine*, vol. 9, no. 1, pp. 14–37, January 1992.
- [19] C. Plapous, C. Marro, L. Mauuary, and P. Scalart, "A two-step noise reduction technique," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, May 2004, pp. 289–292.
- [20] C. Marro, Y. Mahieux, and K. U. Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with post-filtering," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 3, pp. 240–259, May 1998.
- [21] E. R. Ferrara, "Fast implementation of lms adaptive filters," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-28, no. 4, pp. 474–475, August 1980.